

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:47 ; Search time 91.75 Seconds  
(without alignments)  
24.287 Million cell updates/sec

Title: US-09-331-631A-5\_COPY\_76\_144  
Perfect score: 381  
Sequence: 1 NR0RDPQOQYEQCQRCQRR.....EEQGREDEKYEERKMGSDN\_69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues  
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	38.2	588	1	VCLB_GOSHI
2	128	33.6	605	1	VCLB_GOSHI
3	103	27.0	1898	1	TRHY_HUMAN
4	100	26.2	1407	1	TRHY_RABIT
5	97.5	25.6	877	1	INCE_CHICK
6	97	25.5	482	1	U2R2_HUMAN
7	95.5	25.1	1023	1	GLT_DROME
8	95	24.9	524	1	SBP_SOYBN
9	95	24.9	905	1	SNE5_YEAST
10	94	24.7	648	1	KAPC_DICDI
11	94	24.7	1085	1	YAP4_MACFA
12	93	24.4	429	1	APAF_SCHPO
13	93	24.4	2124	1	Y192_HUMAN
14	92.5	24.3	1403	1	PRO_DROME
15	91.5	24.0	573	1	GLB1_MAIZE
16	90	23.6	678	1	GAPR_PLAIF
17	90	23.6	1344	1	IF3A_MOUSE
18	89.5	23.5	758	1	YM38_YEAST
19	89	23.4	285	1	INVO_CANFA
20	89	23.4	1382	1	IF3A_HUMAN
21	88	23.1	1157	1	Y182_HUMAN
22	88	23.0	1905	1	TAGB_DICDI
23	87.5	23.0	538	1	CK1L_YEAST
24	87.5	23.0	1549	1	TRHY_SHEEP
25	85.5	22.4	550	1	BLSA_HUMAN
26	85.5	22.4	572	1	MOES_LYTV
27	85.5	22.4	695	1	XE7_HUMAN
28	84.5	22.2	47	1	AGRP_LUFCY
29	84	22.0	1240	1	YNI1_YEAST
30	83.5	21.9	466	1	CYP8_CAEEL
31	83.5	21.9	471	1	RUI7_XENLA
32	83	21.8	479	1	U2R1_HUMAN
33	82.5	21.7	444	1	CEB_DROME

34	82.5	21.7	708	1	GBF_DICDI	P36417 dictyostell
35	82.5	21.7	743	1	ABRA_PLAFC	P22620 plasmodium
36	82.5	21.7	1048	1	SRA4_RAT	O63627 rattus norv
37	82.5	21.7	1336	1	SEC3_YEAST	P23332 saccharomyc
38	82	21.5	321	1	ABRA_PLAIF	P23746 plasmodium
39	81	21.3	386	1	ARPA_STRPY	P13050 streptococc
40	81	21.3	389	1	M49_STRPY	P16947 streptococc
41	81	21.3	462	1	U2R2_MOUSE	O62377 mus musculu
42	81	21.3	614	1	RUI7_HUMAN	P08621 homo sapien
43	80.5	21.1	338	1	LEGB_PEA	P14594 pisum sativ
44	80.5	21.1	1505	1	SINA_DROME	O24167 drosophil
45	79	20.7	347	1	INVO_PIG	P18175 sus scrofa

## ALIGNMENTS

```

RESULT 1
VCLB_GOSHI STANDARD; PRT; 588 AA.
AC P09801;
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
OC Magnoliophyta: Eudicotyledons, core eudicots: Rosidae, eurosids II;
OC Malvales: Malvaceae: Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XVIII. CDNA and amino acid sequences of the members of
RT the storage protein families."
RL Plant Mol. Biol. 7:475-489(1986).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLICININ, ETC.).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M16891; AAA33071.1; -
CC PIR: A30838; FMCNAB.
CC DR HSSP: P50477; ICAX.
CC DR INTERPRO: IPR001113; -
CC DR PFMW: PF00546; Seedstore_7s; 1.
CC FT SIGNAL 1 25
CC FT CHAIN 26 588 VICILIN C72.
CC SQ SEQUENCE 588 AA: 69729 MW: 63569929AB8ADB CRC64;

```

Query Match 38.2%; Score 145.5; DB 1; Length 588;

Best local Similarity 35.4%; Pred. No. 2e-05; Mismatches 14; Indels 19; Gaps 2;

Matches 28; Conservative 18; Mismatches 14; Indels 19; Gaps 2;

1 NR0RDPQOQYEQCQRCQRRPRHMQICQRCRRYKRRKQKRYEEO-----53

78 HRPEDEQRRYEEQOC--RQDEERORPQOCQRCLEKRFQEQOOSORQFQCOHQOE 135

54 -----REDEKRYEE 62

136 QRPERRQOCVRECRERYOE 154

```

RESULT 2
VICLA_GOSHI STANDARD: PRT: 605 AA.
ID VICLA_GOSHI
AC P09799:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RN SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA3069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -
DR PFM: PF00546; Seedstore_7s; 1.
DR KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
SQ SEQUENCE 605 AA: 71049 MW: C9DB9371C976953B CRC64;

```

```

RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steiner P.M.;
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a coniferyl cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein."
RL J. Biol. Chem. 268:12164-12176(1993).
RN [12]
RN SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE-93315897; PubMed-7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steiner P.M.;
RT "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis."
RL J. Invest. Dermatol. 101:655-715(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L09190; AAA6582.1; -
DR PIR: A45973; A45973.
DR HSSP: P02633; IBOC.
DR MIM: 190370; -.
DR INTERPRO: IPR001751; -
DR INTERPRO: IPR002048; -
DR PFM: PF01023; S.100; 1.
DR PFM: PF00036; eHand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 93
FT CA_BIND 62 73
FT DOMAIN 314 390
FT REPEAT 314 326
FT REPEAT 327 339
FT REPEAT 340 351
FT REPEAT 352 364
FT REPEAT 365 377

```

```

FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-O-O-L.
FT REPEAT 391 396 2-1.
FT REPEAT 397 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 923 1162 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 953 982 4-2.
FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F -> L (IN REF. 2).
FT CONFLICT 1754 1801 Q -> K (IN REF. 2).
FT CONFLICT 1857 1857 V -> G (IN REF. 2).
FT CONFLICT 1880 1880
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

```

Query Match 27.0%; Score 103; DB 1; Length 1898;

Best Local Similarity 31.0%; Pred. No. 0.16; Mismatches 20; Indels 6; Gaps 1;

```

QY 2 RQRPQOQYECOCR-----RCQRETEPRHMOICQRCERYEKKRQKRYEQORE 55
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 267 RQRLQRELEIRERERQRLRERQREERQOQRLRERQRLRERQREERQOQRE 326
QY 56 DEEKYEERAKE 66
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 327 RREQOERREO 337

```

```

RESULT 4
TRHM_RABIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
    IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
    INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
    LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
    ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
    WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
    ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
    ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
    DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
    THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
    THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
    THE EPIDERMIS.

```

```

CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
    CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
    ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
    OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
    ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
    THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
    DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
    THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
    DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
    PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
    FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation-
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC EMBL; Z19092; CAA79519.1; -.
DR PIR; S28589; S28589.
DR HSSP; P02633; 1BOC.
DR INTERPRO; IPR001751; -.
DR INTERPRO; IPR002048; -.
DR PFAM; PF01023; S100; 1.
DR PFAM; PF00036; ehfand; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91 S-100 LIKE.
FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
SQ SEQUENCE 1407 AA; 183781 MW; AE1D2A159F12B7F CRC64;

```

Query Match 26.2%; Score 100; DB 1; Length 1407;

Best Local Similarity 34.7%; Pred. No. 0.21; Mismatches 19; Indels 18; Gaps 2;

```

QY 3 QRPDQOQYECOCRCRCRTEPRHMOICQRCERYEKKRQKQK-----RYEQOR 54
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 256 QOQLRELEIRERERQRLRERQREERQOQRLRERQRLRERQREERQRL 313
QY 55 EDEKYEERAKE 66
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 314 EDERREORLEQ 325

```

```

RESULT 5
INCE_CHICK STANDARD; PRT; 877 AA.
AC P53352;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE INNER CENTROMERE PROTEIN (INCENP).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
    Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-94012983; PubMed-8408220;
RA Mackay A.M., Eckley D.M., Chue C., Earnshaw W.C.;
RT Molecular analysis of the INCENPs (inner centromere proteins):
    separate domains are required for association with microtubules
    during interphase and with the central spindle during anaphase.
    J. Cell Biol. 123:373-385(1993).
CC -1- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY

```

```

CC CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE
CC NETWORK DURING INTERPHASE.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN
CC INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY
CC METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR
CC ARRAYS THAT TRANSPECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.
CC AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS
CC INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN
CC ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY
CC TELEOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE
CC INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; INCENP I AND INCENP II (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: Z25419; CAAB0906.1; -
DR EMBL: Z25420; CAAB0907.1; -
KM Cell division; Microtubules; Coiled coil; Centromere; Mitosis;
FT VARPPLIC 716 733 COILED COIL (POTENTIAL).
FT VARIANT 257 257 MISSING (IN ISOFORM INCENP I).
FT VARIANT 471 471 A -> T.
FT VARIANT 471 471 E -> Q.
SQ SEQUENCE 877 AA: 100940 MW: AFA703149F555352 CRC64;

Query Match 25.6%; Score 97.5; DB 1; Length 877;
Best Local Similarity 33.3%; Pred. No. 0.22;
Matches 22; Conservative 16; Mismatches 25; Indels 3; Gaps 1;

OY 4 RDPPOOYOCRCORRRETEPRHMOICQRCRRYERKEKROOKRYEORDE--EKY 60
DB 606 RRRODEAKKQALQOEERBRHKLMOKKREDEBRACKLAEQCAEERKEKLAERE 665
OY 61 ERRAKE 66
DB 666 QERKKE 671

RESULT 6
UZR2_HUMAN STANDARD: PRT; 482 AA.
ID UZR2_HUMAN Q15696;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN AUXILIARY FACTOR 35 KDA SUBUNIT
DE RELATED-PROTEIN 2.
GN UZAF1-RS2 OR UZAF1RS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT Uzafl-rs1.";
RL Genomics 30:257-263(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: TO MAMMALIAN SPLICING FACTOR UZAF 35 KDA SUBUNIT.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: D49677; BAA08533.1; -
DR MIM: 300028; -
DR INTERPRO: IPR000504; -
DR INTERPRO: IPR000571; -
DR PFAM: PF00076; rtm; 1.
DR PFAM: PF00642; zf-CCCH; 1.
DR PROSITE: PS50102; RRM; 1.
KM Nuclear protein; Ribonucleoprotein.
FT DOMAIN 46 49 POLY-GLU.
FT DOMAIN 118 123 POLY-GLU.
SQ SEQUENCE 482 AA: 58044 MW: 1DACCB8AC4727A6 CRC64;

Query Match 25.5%; Score 97; DB 1; Length 482;
Best Local Similarity 29.7%; Pred. No. 0.14;
Matches 19; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

OY 3 ORDPPOOYOCRCORRRETEPRHMOICQRCRRYERKEKROOKRYEORDEKYE 62
DB 65 ERORRLHEMLRLRQKQAEFRIRKKEAEAKKROEDERLKLQWEOQRKEEEO 124
OY 63 RKKE 66
DB 125 KRQE 128

RESULT 7
GLT_DROME STANDARD: PRT; 1023 AA.
ID GLT_DROME P33438;
AC 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLUTACTIN PRECURSOR.
DE GLT.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=OREGON-R;
RX MEDLINE=90214632; PubMed=2108864;
RA Olson P.F., Fessler L.I., Nelson R.E., Sterne R.E., Campbell A.G.,
RA Fessler J.H.;
RT "Glutactin, a novel Drosophila basement membrane-related glycoprotein
RT with sequence similarity to serine esterases.";
RL EMBO J. 9:1219-1227(1990).
CC -1- FUNCTION: NOT KNOWN. BINDS CALCIUM IONS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED AND ALSO N-GLYCOSYLATED.
CC -1- PTM: FOUR TYROSINES ARE SULFATED.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-B
CC CARBOXYLESTERASE/LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: X53286; CAA37380.1; -

```



```

DR EMBL: M36482; AAA35062.1; -
DR EMBL: X76053; CAA33652.1; -
DR EMBL: Z36158; CAA85254.1; -
DR PIR: S44551; RGYBS5.
DR PIR: S39145; S39145.
DR SGD: S0000493; SNF5.
KM Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 31 270 GLN-RICH.
FT DOMAIN 72 132 PRO-RICH.
FT DOMAIN 272 324 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 489 588 PRO-RICH.
FT DOMAIN 714 882 ARG/LYS-RICH (BASIC).
FT DOMAIN 755 798 E -> D (IN REF. 1).
FT CONFLICT 564 564
SQ SEQUENCE 905 AA; 102557 MW; A287B4A648D1A35 CRC64;

Query Match
Best Local Similarity 24.9%; Score 95; DB 1; Length 905;
Matches 15; Conservative 30; Mismatches 15; Indels 2; Gaps 1;

QY 7 QQQYEQCQKRCQRETEP--RHMQICQRCERREYKRRKQKRYEQQREDEKYEEM 64
DB 193 QQQQQQLRNQIQRRQQQQRHNNHQIQQQQKQQQQQQQQQQQQQQQQQQQQQ 252
QY 65 KE 66
DB 253 QQ 254

RESULT 10
KAPC_DICD1 STANDARD; PRT; 648 AA.
AC P34099;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT (EC 2.7.1.37).
GN PKAC OR PK2 OR PK3.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91323730; PubMed-1864510;
RA Buerki E., Anjard C., Scholder J.-C., Raymond C.D.;
RT "Isolation of two genes encoding putative protein kinases regulated during Dictyostelium discoideum development.";
RL Gene 102:57-65(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE-93385090; PubMed-8373760;
RA Anjard C., Etcheberry L., Pinaud S., Veron M., Raymond C.D.;
RT "An unusual catalytic subunit for the CAMP-dependent protein kinase of Dictyostelium discoideum.";
RL Biochemistry 32:9532-9538(1993).
RN [3]
RP CHARACTERIZATION.
RX STRAIN-AX3;
RA Mann S.K.O., Yonemoto W.M., Taylor S.S., Firtel R.A.;
RT "Ddck3, which plays essential roles during Dictyostelium development, encodes the catalytic subunit of CAMP-dependent protein kinase.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10701-10705(1992).
CC -1- FUNCTION: ESSENTIAL FOR DIFFERENTIATION AND FRUIT MORPHOGENESIS.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN -> ADP + A PHOSPHOPROTEIN.
CC -1- SUBUNIT: IN DICTYOSTELIUM THE Holoenzyme is a dimer composed of A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC -1- DEVELOPMENTAL STAGE: CAMP ACTIVITY IS LOW IN VEGETATIVELY GROWING AMOEBAE, INCREASES DURING DEVELOPMENT OF AGGREGATION AND REACHES A MAXIMUM AT CUMINATION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CAMP SUBFAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M38703; -; NOT_ANNOTATED_CDS.
DR PIR: J01150; J01150.
DR HSSP: P05132; ZCRK.
DR DICTYDB: D002030; PKAC.
DR INTERPRO: IPR000719; -.
DR INTERPRO: IPR000961; -.
DR INTERPRO: IPR002290; -.
DR PFAM: PF00069; pkinase.1.
DR PFAM: PF00433; pkinase.C.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
KM Transferrase; Serine/threonine-protein kinase; ATP-binding; CAMP;
KW Phosphorylation.
FT DOMAIN 58 64 ASN-RICH.
FT DOMAIN 136 223 GLN-RICH.
FT DOMAIN 233 250 THR-RICH.
FT DOMAIN 336 590 PROTEIN KINASE.
FT NP_BIND 342 350 ATP (BY SIMILARITY).
FT BINDING 365 365 ATP (BY SIMILARITY).
FT ACT_SITE 459 459 BY SIMILARITY.
FT MOD_RES 480 490 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 648 AA; 74458 MW; D0F9B3A48C5D084 CRC64;

Query Match
Best Local Similarity 23.9%; Score 94; DB 1; Length 648;
Matches 17; Conservative 29; Mismatches 19; Indels 6; Gaps 1;

QY 2 RQRDPQQYEQCQKRCQRETEPRHMQICQRCERREYKRRK-----QQRREEQRE 55
DB 144 QQQQQPQQQQPQQQQQQPQQQQQQPQQQQQQPQQQQQQQQQQQQQQQQQQQQQ 203
QY 56 DEEKYEEMKE 66
DB 204 QQQQQQQQQQQQ 214

RESULT 11
YAPA_SCHPO STANDARD; PRT; 1085 AA.
AC Q09863;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 122.9 KDA PROTEIN C29E6.10C IN CHROMOSOME I.
GN SPAC29E6.10C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
CC Schizosaccharomycetaceae; Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-972;
RA Jones L., Murphy L., McNeil A., Simpson I., Harris D., Barrell B.G.,
RA Ralston M.A., Walsh S.V.;
RT Submitted (OCF-1995) to the EMBL/Genbank/DBJ databases.
RL Submitted (OCF-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YNL091W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```
SO SEQUENCE 2124 AA; 237207 MW; 255FB9419EC39FA2 CRC64;
```

Query Match            24.4%; Score 93; DB 1; Length 2124;  
Best Local Similarity     21.7%; Pred. No. 1.1;  
Matches       15; Conservative      35; Mismatches      15; Indels          4; Gaps            1

OY            2 RQRDQOQQVEQCQRKRCORRETEPRHMOICQ-----RCEERYEKERKKOCKRYEQGRDE 57  
             ::::|||||::::::::::|||:|:::|||||:::|||||:|  
Db       2001 QOOQOQOQOQOQOQOQOQOQOQHHRQOQOQLRLROQOQOQOQOQOQOQOQOQOQOQO 2060  
  
OY            58 EKYERRMKE 66  
             :::::::::::  
Db       2061 QQHQOQOQOQ 2069

RESULT       14  
PRO\_DROME                ;  
ID PRO\_DROME           STANDARD; PRT: 1403 AA.

AC       p29617,  
DT       01-APR-1993 (Rel. 25, Created)  
DT       01-FEB-1996 (Rel. 33, Last sequence update)  
DT       30-MAY-2000 (Rel. 39, Last annotation update)  
DE       PROTEIN PROSPERO.  
CN       Drosophila melanogaster (Fruit fly).  
OS       Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC       Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
MN       Epiphytoidea; Drosophilidae; Drosophila.  
[1]  
RP       SEQUENCE FROM N.A.  
RX       MEDLINE=92069760; PubMed=1720353;  
RA       Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;  
RT       "Prospero is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in Drosophila."  
RL       Cell 67:941-953(1991).

[2]  
RN       SEQUENCE FROM N.A.  
RP       MEDLINE=92171948; PubMed=1540176;  
RX       Matsuzaki F., Kolzumi K., Hama C., Yoshioka T., Nabeshima Y.;  
RA       "Cloning of the Drosophila prospero gene and its expression in ganglion mother cells."  
RT       Biochem. Biophys. Res. Commun. 182:1326-1332(1992).

[3]  
RN       SEQUENCE FROM N.A.  
RP       MEDLINE=93083413; PubMed=1842358;  
RX       Chn-Lagraff O., Wright D.M., McNeil L.K., Doe C.Q.;  
RT       "The prospero gene encodes a divergent homeodomain protein that controls neuronal identity in Drosophila.";  
RL       Development Suppl. 2:79-85(1991).

[4]  
RN       SIMILARITY TO C.ELEGANS CEH-26.  
RP       Medline=94212446; PubMed=7909177;  
RX       Buerklin T.R.;  
RA       "A Caenorhabditis elegans prospero homologue defines a novel domain.", Trends Blochm. Sci. 19:70-71(1994).

-I- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE TRANSCRIPTION BY BINDING TO DNA.

-I- SUBCELLULAR LOCATION: NUCLEAR.

-I- ALTERNATIVE PRODUCTS: 2 ISOFORMS, PROS-L (SHOWN HERE) AND PROS-TERMIAL AMINO ACIDS OF THE HOMEDOMAIN.

-I- TISSUE SPECIFICITY: NEURONAL PRECURSORS, EXPRESSED IN THE DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.

-I- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING FORMATION.

-I- SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEBOX PROTEINS.

```
CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M81389; AAA28841.1; -.
CC DR EMBL; D10609; BAA01464.1; -.
CC DR EMBL; Z11743; CAA77802.1; -.
CC PIR; A41089; A41089.
CC PIR; JQ1397; JQ1397.
CC FLYBASE; FBgn0004595; pros.
CC KW Nuclear protein; Transcription regulation; DNA-binding; Homeobox; Developmental protein; Alternative splicing.
CC FT DOMAIN 4 12 POLY-SER.
CC FT DOMAIN 28 31 POLY-SER.
CC FT DOMAIN 32 35 POLY-ASN.
CC FT DOMAIN 188 191 POLY-ALA.
CC FT DOMAIN 253 260 POLY-GLN.
CC FT DOMAIN 270 276 POLY-ASN.
CC FT DOMAIN 282 286 POLY-ASN.
CC FT DOMAIN 431 437 POLY-ASP.
CC FT DOMAIN 505 508 POLY-ALA.
CC FT DOMAIN 717 737 POLY-GLN.
CC FT DOMAIN 754 763 POLY-GLN.
CC FT DOMAIN 766 772 POLY-GLN.
CC FT DOMAIN 934 937 POLY-ALA.
CC FT DOMAIN 952 957 POLY-GLN.
CC FT DOMAIN 960 963 POLY-GLN.
CC FT DOMAIN 966 970 POLY-GLN.
CC FT DOMAIN 991 998 NUCLEAR LOCALIZATION SIGNAL.
CC FT DOMAIN 1029 1048 POLY-GLN.
CC FT DOMAIN 1074 1082 POLY-PRO.
CC FT DNABIND 1241 1303 HOMEBOX (ATYPICAL).
CC FT DOMAIN 1304 1403 PROSPERO-LIKE.
CC FT VARSPLIC 1216 1244 MISSING (IN ISOFORM PROS-S).
CC FT CONFLICT 76 98 AKM1NELFGROMKOADTSGLP -> GODAERAWVPDEA GPGHNPMA (IN REF. 1).
CC FT CONFLICT 120 144 IGSLSNSTSKLLQQOHNNNSIAPANS -> NLALDFHYVAAA AAATITALLPLIG (IN REF. 1).
CC FT CONFLICT 418 418 H -> Q (IN REF. 2).
CC FT CONFLICT 677 677 A -> C (IN REF. 1).
CC FT CONFLICT 802 802 A -> S (IN REF. 1).
CC FT CONFLICT 958 958 T -> S (IN REF. 1).
CC FT CONFLICT 1048 1048 Q -> QDDNQ (IN REF. 1).
CC SO SEQUENCE 1403 AA; 153569 MW; 9EFB9973E24E23BE CRC64;
CC
CC Query Match 24.3%; Score 92.5; DB 1; Length 1403;
CC Best Local Similarity 33.3%; Pred. No. 0.85;
CC Matches 19; Conservative 19; Mismatches 10; Indels 9; Gaps 1;
CC
CC QY 5 DPQQDYEQCKRCORRETEPRHMICQRCEERYEKRRKQOKRAYEEQOREDEKEYE 61
CC : ||| : |:: :||:
CC Db 716 EQDQDQDQDQDQDQDQDQDQ-----QQEQDRRFQEQDQEQRRRKEQDQDQIQDQDQ 763
CC
CC RESULT 15
CC ID GBL1_MAIZE STANDARD; PRT; 573 AA.
CC AC P15590;
CC DT 01-APR-1990 (Rel. 14, Created)
CC DT 01-AUG-1990 (Rel. 15, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE GLOBULIN-1 S ALLELE PRECURSOR (GBL1-S) (7S-LIKE).
CC GN GBL1.
CC OS Zea mays (Maize).
CC CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
CC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
CC [1]
```



```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. INBERD LINE VA26;
RA Belanger F.C., Kriz A.L.;
RT "Molecular characterization of the major maize embryo globulin encoded
RL by the G1b1 gene."
RL Plant Physiol. 91:636-643(1989).
RN [2]
RP SEQUENCE OF 87-100.
RX MEDLINE=69374022; Pubmed=2775172;
RA Kriz A.L.;
RT "Characterization of embryo globulins encoded by the maize G1b
RL genes."
RL Biochem. Genet. 27:239-251(1989).
CC -|- PTM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
CC MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
CC -|- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLB1 ALLELES HAVE
CC THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
CC PROTEINS, RESPECTIVELY.
CC -|- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24845; AAA33467.1; -.
DR HSSP: P50477; ICAM.
DR MAIZEDB: 30181; -.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KM Seed storage protein: Signal. OR 21 (POTENTIAL).
FT STGNL 1 18
FT PROPE 19 86
FT CHAIN 87 573 GLOBULIN-1 S ALLELE.
FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 573 AA; 65029 MW; 525ED1D00A062976 CRC64;

```

```

Query Match 24.08; Score 91.5; DB 1; Length 573;
Best Local Similarity 33.38; Pred. No. 0.47;
Matches 22; Conservative 16; Mismatches 13; Indels 15; Gaps 4;

```

```

QY 12 QCGKRCORRETPRHMQ-IGQRCERYEKEKRRKQKRYE-----EQGREDEPKY 60
DB 39 KCVKRCEDR--FWHQRPLEQC-REEREKREKRSRHADDRSGSGSSEDEREREQK 94
QY 61 EERNKKE 66
DB 95 EKKOKD 100

```

Search completed: March 1, 2001, 16:16:49  
 Job time: 429 sec

